

## EXHIBIT A

(12) **United States Patent**  
**Moricz et al.**

(10) **Patent No.:** **US 7,693,956 B2**  
(45) **Date of Patent:** **\*Apr. 6, 2010**

(54) **ACCESSING DEEP WEB INFORMATION  
USING A SEARCH ENGINE**

(75) Inventors: **Michael Z. Moricz**, 1575 Tenaka Pl.,  
Apt. H8, Sunnyvale, CA (US) 94087;  
**John R. Kasha, Jr.**, Falls Church, VA  
(US)

(73) Assignee: **Michael Z. Moricz**, Bellevue, WA (US)

(\*) Notice: Subject to any disclaimer, the term of this  
patent is extended or adjusted under 35  
U.S.C. 154(b) by 2123 days.

This patent is subject to a terminal dis-  
claimer.

(21) Appl. No.: **10/397,992**

(22) Filed: **Mar. 25, 2003**

(65) **Prior Publication Data**

US 2003/0212737 A1 Nov. 13, 2003

**Related U.S. Application Data**

(60) Provisional application No. 60/366,817, filed on Mar.  
25, 2002.

(51) **Int. Cl.**  
**G06F 15/16** (2006.01)

(52) **U.S. Cl.** ..... **709/217; 709/203; 707/3;**  
**707/10; 705/14**

(58) **Field of Classification Search** ..... **709/217;**  
**709/203; 707/10, 102, 7, 104.1, 9, 5, 3, 4;**  
**705/14**

See application file for complete search history.

(56) **References Cited**

**U.S. PATENT DOCUMENTS**

5,864,863 A 1/1999 Burrows  
6,278,993 B1 8/2001 Kumar et al.  
6,321,251 B1 11/2001 Deisinger et al.  
6,487,555 B1 11/2002 Bharat et al.

6,718,365 B1 \* 4/2004 Dutta ..... 709/217  
6,993,569 B1 \* 1/2006 Rees, Jr. .... 709/217  
7,277,923 B2 \* 10/2007 Rensin et al. .... 709/217  
2001/0049676 A1 \* 12/2001 Kepler et al. .... 707/3  
2001/0049687 A1 \* 12/2001 Russell ..... 707/3  
2002/0123991 A1 \* 9/2002 Asami et al. .... 707/3  
2002/0133381 A1 \* 9/2002 Tso ..... 705/5

(Continued)

**OTHER PUBLICATIONS**

IncyWincy, webpages, copyright 2001.\*

(Continued)

*Primary Examiner*—Ario Etienne

*Assistant Examiner*—Barbara N Burgess

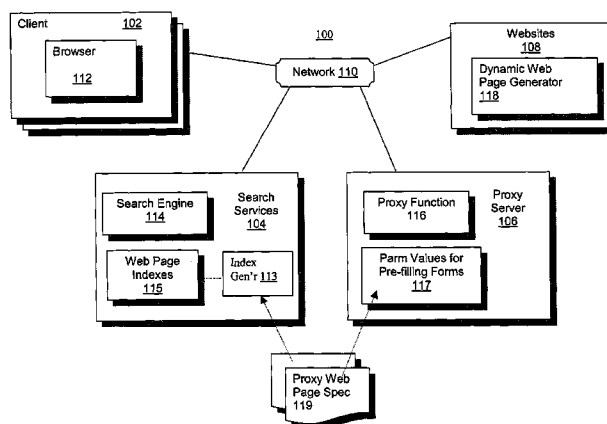
(74) *Attorney, Agent, or Firm*—Schwabe, Williamson &  
Wyatt P.C.

(57)

**ABSTRACT**

Associated query parameters of queries of query answer web pages of a plurality of websites corresponding to a set of search criteria are identified and reconciled. Proxy web page specifications comprising the queries and their associated query parameters are created, and submitted for processing by a search engine to create a plurality of indices to effectively index the dynamic query answer web pages. For associated query parameters to be posted to the websites, the associated query parameters for pre-filling the forms are pre-stored at a proxy server, and the corresponding queries included in the indices, modified to address the proxy servers. During operation, on receipt of the search criteria from a client, the search engine returns a search answer web page including the indices that effectively indexes the dynamic query answer web pages. In alternate embodiment, the effective indices may be automatically selected

**10 Claims, 22 Drawing Sheets**



**US 7,693,956 B2**

Page 2

## U.S. PATENT DOCUMENTS

2002/0152101 A1 \* 10/2002 Lawson et al. .... 705/6  
 2003/0088544 A1 5/2003 Kan et al.

## OTHER PUBLICATIONS

IncyWincy, Background Information, copyright 2001.\*

IncyWincy, Article, "Loop Improvements Launches IncyWincy, the Invisible Web Search Engine", Aug. 16, 2000.\*

Warnick, W. et al., "Searching the Deep Web: Directed Query Engine Applications at the Department of Energy", Jan. 1, 2001, <<http://www.dlib.org/dlib/january01/warnick/01warnick.html>> [retrieved Jul. 19, 2006].

"Quigo Unveils Deep Web Search Technologies", Aug. 15, 2001, <<http://www.quigo.com/press081501.htm>> [retrieved Jul. 19, 2006].

Bruemmer, P., "Make Way for the Deep Crawl", Aug. 29, 2001, <<http://www.clickz.com/experts/search/opt/article.php/874131>> [retrieved Jul. 19, 2006].

Raghavan, S. and Garcia-Molina, H., "Crawling the Hidden Web" Sep. 11, 2001, Proceedings of 27th Int'l Conf. on VLDB, pp. 129-139, <[http://www.dia.uniroma3.it/~vldbproc/017\\_129.pdf](http://www.dia.uniroma3.it/~vldbproc/017_129.pdf)> [retrieved Jul. 19, 2006].

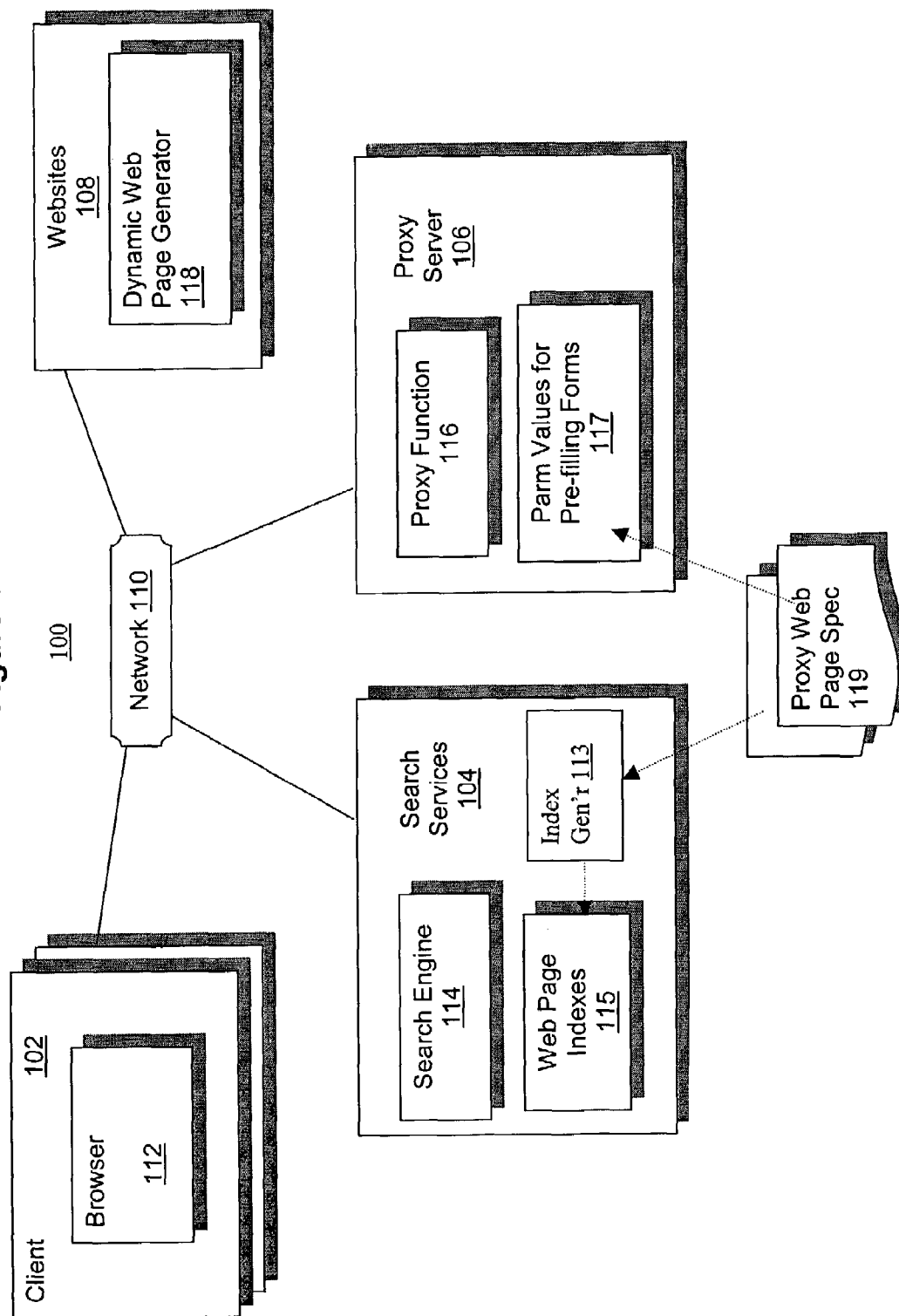
Bergman, M., "The Deep Web: Surfacing Hidden Value" Sep. 24, 2001, Journal of Electronic Publishing 7(1), <<http://www.brightplanet.com/pdf/deepwebwhitepaper.pdf>> [retrieved Jul. 19, 2006].

Bruemmer, P., "Indexing Deep Web Content", Mar. 27, 2002, <[http://www.searchengineguide.com/wi/2002/0327\\_wi2.html](http://www.searchengineguide.com/wi/2002/0327_wi2.html)> [retrieved Jul. 19, 2006].

Mills, E., "Google flight search takes off", Oct. 28, 2005, <[http://www.news.com/Google+flight+search+takes+off/2100-1038\\_3-5917821.html](http://www.news.com/Google+flight+search+takes+off/2100-1038_3-5917821.html)> [retrieved Jul. 19, 2006].

\* cited by examiner

Figure 1

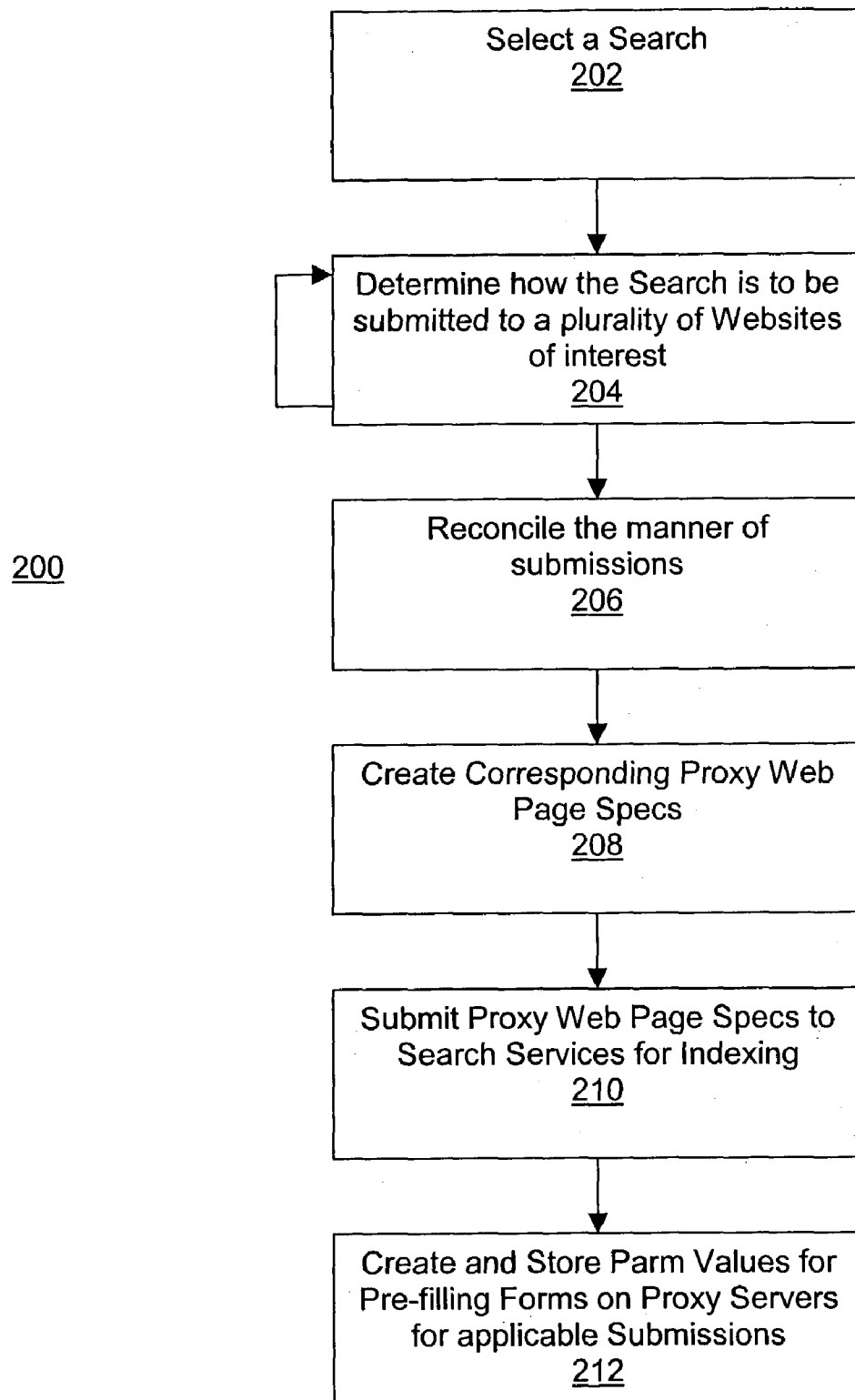


**U.S. Patent**

**Apr. 6, 2010**

**Sheet 2 of 22**

**US 7,693,956 B2**



**Figure 2**

**U.S. Patent****Apr. 6, 2010****Sheet 3 of 22****US 7,693,956 B2**

Uniform Resource Locators <u>302</u>	Search Criteria <u>304</u>
<a href="http://www.automall.com/used.html?make=ford&amp;model=focus&amp;zip=22043">http://www.automall.com/used.html?make=ford&amp;model=focus&amp;zip=22043</a>	Focus 22043
<a href="http://www.automall.com/used.html?make=ford&amp;model=focus&amp;zip=22043">http://www.automall.com/used.html?make=ford&amp;model=focus&amp;zip=22043</a>	22043 Focus
<a href="http://www.automall.com/used.html?make=ford&amp;model=focus&amp;zip=22043">http://www.automall.com/used.html?make=ford&amp;model=focus&amp;zip=22043</a>	Focus Falls Church
<a href="http://www.automall.com/used.html?make=ford&amp;model=focus&amp;zip=22043">http://www.automall.com/used.html?make=ford&amp;model=focus&amp;zip=22043</a>	Falls Church Focus
<a href="http://www.automall.com/used.html?make=ford&amp;model=taurus&amp;zip=22043">http://www.automall.com/used.html?make=ford&amp;model=taurus&amp;zip=22043</a>	Taurus 22043
<a href="http://www.automall.com/used.html?make=ford&amp;model=taurus&amp;zip=22043">http://www.automall.com/used.html?make=ford&amp;model=taurus&amp;zip=22043</a>	22043 Taurus
<a href="http://www.automall.com/used.html?make=ford&amp;model=taurus&amp;zip=22043">http://www.automall.com/used.html?make=ford&amp;model=taurus&amp;zip=22043</a>	Taurus Falls Church
<a href="http://www.automall.com/used.html?make=ford&amp;model=taurus&amp;zip=22043">http://www.automall.com/used.html?make=ford&amp;model=taurus&amp;zip=22043</a>	Falls Church Taurus
<a href="http://www.oursite.com/proxyjoescars.html?Make=ford&amp;Model=focus&amp;Zip=22043&amp;Year=1982-2002">http://www.oursite.com/proxyjoescars.html?Make=ford&amp;Model=focus&amp;Zip=22043&amp;Year=1982-2002</a>	Focus 22043
<a href="http://www.oursite.com/proxyjoescars.html?Make=ford&amp;Model=focus&amp;Zip=22043&amp;Year=1982-2002">http://www.oursite.com/proxyjoescars.html?Make=ford&amp;Model=focus&amp;Zip=22043&amp;Year=1982-2002</a>	22043 Focus
<a href="http://www.oursite.com/proxyjoescars.html?Make=ford&amp;Model=focus&amp;Zip=22043&amp;Year=1982-2002">http://www.oursite.com/proxyjoescars.html?Make=ford&amp;Model=focus&amp;Zip=22043&amp;Year=1982-2002</a>	Focus Falls Church
<a href="http://www.oursite.com/proxyjoescars.html?Make=ford&amp;Model=focus&amp;Zip=&amp;22043&amp;Year=1982-2002">http://www.oursite.com/proxyjoescars.html?Make=ford&amp;Model=focus&amp;Zip=&amp;22043&amp;Year=1982-2002</a>	Falls Church Focus
<a href="http://www.oursite.com/proxyjoescars.html?Make=ford&amp;Model=granada&amp;Zip=22043&amp;Year=1982-2002">http://www.oursite.com/proxyjoescars.html?Make=ford&amp;Model=granada&amp;Zip=22043&amp;Year=1982-2002</a>	Granada 22043
<a href="http://www.oursite.com/proxyjoescars.html?Make=ford&amp;Model=granada&amp;Zip=22043&amp;Year=1982-2002">http://www.oursite.com/proxyjoescars.html?Make=ford&amp;Model=granada&amp;Zip=22043&amp;Year=1982-2002</a>	22043 Granada
<a href="http://www.oursite.com/proxyjoescars.html?Make=ford&amp;Model=granada&amp;Zip=22043&amp;Year=1982-2002">http://www.oursite.com/proxyjoescars.html?Make=ford&amp;Model=granada&amp;Zip=22043&amp;Year=1982-2002</a>	Granada Falls Church
<a href="http://www.oursite.com/proxyjoescars.html?Make=ford&amp;Model=granada&amp;Zip=22043&amp;Year=1982-2002">http://www.oursite.com/proxyjoescars.html?Make=ford&amp;Model=granada&amp;Zip=22043&amp;Year=1982-2002</a>	Falls Church Granada

**Figure 3a**

**U.S. Patent****Apr. 6, 2010****Sheet 4 of 22****US 7,693,956 B2**

Uniform Resource Locators <u>312</u>	Search Criteria <u>314</u>
<a href="http://www.generaltravel.com/res.html?persons=1&amp;from=bwi&amp;to=sfo&amp;leave=0810&amp;return=0815">http://www.generaltravel.com/res.html?persons=1&amp;from=bwi&amp;to=sfo&amp;leave=0810&amp;return=0815</a>	Baltimore San Francisco 0810 0815
<a href="http://www.generaltravel.com/res.html?persons=1&amp;from=bwi&amp;to=sfo&amp;leave=0810&amp;return=0815">http://www.generaltravel.com/res.html?persons=1&amp;from=bwi&amp;to=sfo&amp;leave=0810&amp;return=0815</a>	BWI San Francisco 0810 0815
<a href="http://www.generaltravel.com/res.html?persons=1&amp;from=bwi&amp;to=sfo&amp;leave=0810&amp;return=0815">http://www.generaltravel.com/res.html?persons=1&amp;from=bwi&amp;to=sfo&amp;leave=0810&amp;return=0815</a>	Baltimore SFO 0810 0815
<a href="http://www.generaltravel.com/res.html?persons=1&amp;from=bwi&amp;to=sfo&amp;leave=0810&amp;return=0815">http://www.generaltravel.com/res.html?persons=1&amp;from=bwi&amp;to=sfo&amp;leave=0810&amp;return=0815</a>	BWI SFO 0810 0815

**Figure 3b**

119a

```
<SEARCHFORMAT>
<DOC
url=http://www.automall.com/used.html?make=ford&model=focus&zip=22043>
<CONTENT type="text/html">
<HTML>
<HEAD>
<TITLE>Ford Focus 22043 - Automall.com</TITLE>
<META name="keywords" content="Focus, 22043">
</HEAD>
<BODY>
<H1> Ford Focus 22043 - Automall.com</H1>
<P>
<H2>Find used car pricing information for Focus 22043 at Automall.com.</H2>
<H2>Automall.com is your source for used car information.</H2>
</BODY>
</HTML>
</CONTENT>
</DOC>
</SEARCHFORMAT>
```

**Figure 4a**



119b

```
<SEARCHFORMAT>
<DOC
url=http://www.generaltravel.com/res.html?persons=1&from=bwi&to=sfo&leave=08
10&return=0815>
<CONTENT type="text/html">
<HTML>
<HEAD>
<TITLE>Leave BWI 08/10/02, Return from SFO 08/15/02 –
GeneralTravel.com</TITLE>
<META name="keywords" content="Baltimore,San Francisco,0810,0815">
</HEAD>
<BODY>
<H1> Leave BWI 08/10/02, Return from SFO 08/15/02 – GeneralTravel.com </H1>
<P>
<H2>Find airline reservation information for "Baltimore San Francisco 0810 0815"
at GeneralTravel.com</H2>
<H2>GeneralTravel.com can make all your travel reservations.</H2>
</BODY>
</HTML>
</CONTENT>
</DOC>
</SEARCHFORMAT>
```

Figure 4b

500a

Ford Focus 22043 - Automall.com ~ 502a

...Find used car pricing information for Focus 22043 at Automall.com

Automall.com is your source for used car information...

<http://www.automall.com/used.html?make=ford&model=focus&zip=22043>

Ford Focus 22043 1982-2002 - JoesCars.com ~ 502b

...Find used car pricing information for Focus 22043 at JoesCars.com.

JoesCars.com is a cool place to find used car information...

<http://www.oursite.com/proxyjoescars.html?Make=ford&Model=focus&Zip=22043&Year=1982-2002>

**Figure 5a**

**U.S. Patent**

**Apr. 6, 2010**

**Sheet 8 of 22**

**US 7,693,956 B2**

500b

Leave BWI 08/10/02, Return from SFO 08/15/02 – GeneralTravel.com ~ 502c

...Find airline reservation information for "0810 Baltimore San Francisco 0815" at GeneralTravel.com

GeneralTravel.com can make all your travel reservations...

<http://www.generaltravel.com/res.html?persons=1&from=bwi&to=sfo&leave=0810&return=0815>

Leave BWI 08/10/02, Return from SFO 08/15/02 – FastJet.com ~ 502d

...Find airline reservation information for "0810 Baltimore San Francisco 0815" at FastJet.com

The quickest and best service in the airline industry...

<http://www.fastjet.com/res.html?fdest=bwi&tdest=sfo&start=0810&end=0815>

**Figure 5b**

117a

```
<form name="home" method="POST" action="used.html">
  <b>Used Car Search<br>
  Get FREE listings in your area.</b><br>
  <b>Make:</b><br>
    <select name="used_make">
      <option value="">Select A Make</option>
      <option value='Acura'>Acura</option>
      <option value='Buick'>Buick</option>
      <option value='Chrysler'>Chrysler</option>
      <option value='Daewoo'>Daewoo</option>
      <option selected value='Ford'>Ford</option>

    </select>
  <b>Model:</b><br>
    <select name="used_model">
      <option value="">Select A Model</option>
      <option value="Astro">Astro</option>
      <option value="Avalanche 1500">Avalanche 1500</option>

      <option selected value="Focus">Focus</option>

    </select>
  <b>Your ZIP</b><br>
    <input type="text" name="zipcode" value="22043" size="8" maxlength="5">
    <br>
    <input type="Submit" name="cmdSubmit" value="Continue">
</form>
```

Figure 6a

**U.S. Patent**

**Apr. 6, 2010**

**Sheet 10 of 22**

**US 7,693,956 B2**

602a



The image shows a screenshot of a web form titled "Used Car Search". Below the title is the text "Get FREE listings in your area." The form contains several input fields and dropdown menus. The "Make:" field has a dropdown menu with "Ford" selected. The "Model:" field has a dropdown menu with "Chevrolet" selected, and a "Focus" option is visible below it. The "Your ZIP:" field contains the text "22043". At the bottom of the form is a button labeled "Continue >>".

**Used Car Search**  
Get FREE listings in your area.

Make:  
Ford

Model:  
Chevrolet  
Focus

Your ZIP:  
22043

**Continue >>**

**Figure 6b**

117b

```

<form method="POST" action="res.html" name="FlightSearch">
  <b>Fare Finder</b>
  <a href="more_options.html">More options</a><br>
  From <input type="text" name="from" size=11 value="bwi" >
  To  <input type="text" name="to" size=11 value="sfo" >

  <input type="radio" name="Way" value="RoundTrip" checked >Round trip
  <input type="radio" name="Way" value="OneWay" >One way<br>
  <a href="multi_city.html">Multi city</a><br>

  Persons <select name="persons">
    <option value="1" selected>1</option>

  </select>

  Departure date<br>
  <select name="leave_month">

    <option value="05" selected>May</option>

  </select>
  <select name="leave_date">

    <option value="04" selected>4</option>
    <option value="05" >5</option>

  </select>
  <select name="leave_td">
    <option value="morning" selected>Morning</option>

  </select>
  <br>

Return portion omitted



  Search by
  <input type="radio" name="lowestfares" value="schedule">Schedule
  <input type="radio" name="lowestfares" value="price" checked >Price
  <br>
  <input type="submit" value="Check" >
</form>

```

Figure 6c


U.S. Patent

Apr. 6, 2010

Sheet 12 of 22

US 7,693,956 B2

602b

 **Fare finder** [More options](#)

From

bwi

To

sfo


☒ Round trip ☐ One way [Multi city](#)

Departure date

May ▼

4 ▼

Morning ▼




Return

May ▼

6 ▼

Morning ▼



Search by ☐ Schedule ☒ Price

Passengers 

1 ▼

Check

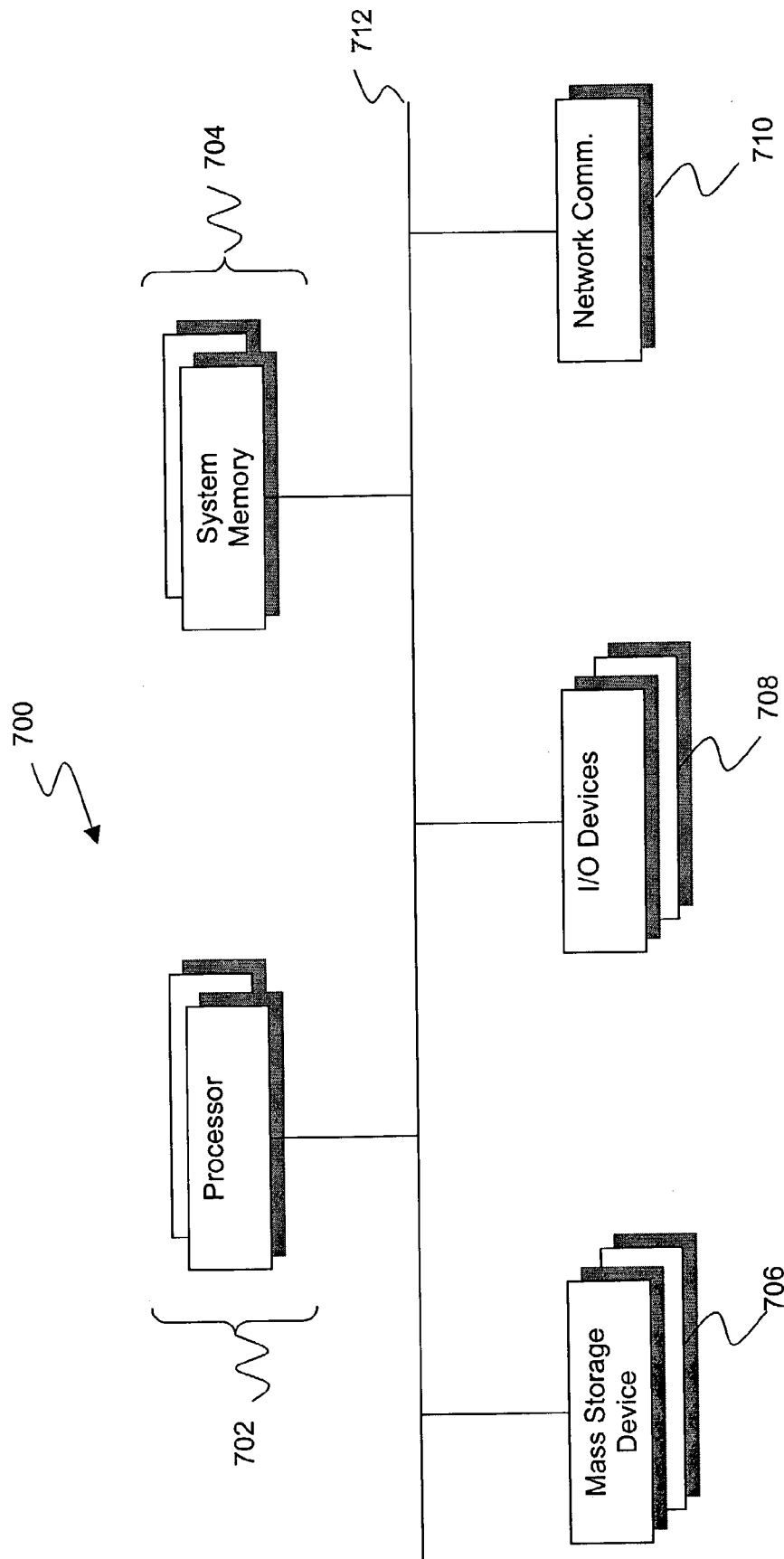
Figure 6d

**U.S. Patent**

**Apr. 6, 2010**

**Sheet 13 of 22**

**US 7,693,956 B2**



**Figure 7**



U.S. Patent

Apr. 6, 2010

Sheet 14 of 22

US 7,693,956 B2

<p><u>PRICELINE -- Name Your Own Price <b>FLIGHT DEALS!</b></u>  <a href="http://www.priceline.com">www.priceline.com</a> Bid now for great deals. Get an answer in as little as 2 minutes.</p> <p><u><b>FLIGHT Deals - Save on AIRLINE TICKETS Here!</b></u>  <a href="http://www.Orbitz.com">www.Orbitz.com</a> <b>FLIGHT</b> - Compare FARES on 450+ Airlines - Book NOW!</p> <p><u>Horizon Air Lowers Fares from Fresno to Portland and Seattle</u>  ...advance from Fresno to Portland or <b>Seattle</b> is \$105... and are nonrefundable; however,  the <b>flight</b> ticket can... to Sacramento, and from Portland and <b>San Jose</b> to Tucson...  <a href="http://www.alaskasworld.com/newsroom/qxnews/QXstories/QX_20020319_145647.asp">www.alaskasworld.com/newsroom/qxnews/QXstories/QX_20020319_145647.asp</a> - 13k - <u>Cached</u> - <u>Similar pages</u></p> <p><u>Southwest Airlines Vacations - Vacation travel packages to Las...</u>  ...Sacramento, San Jose, Oakland, from \$259, Seattle, from \$359... BAY AREA Sample two-night  San Francisco Bay... Flight service for San Francisco Bay area packages is...  <a href="http://www.swavacations.com/about/press/calsale1.asp">www.swavacations.com/about/press/calsale1.asp</a> - 31k - <u>Cached</u> - <u>Similar pages</u></p>	Sponsored Link
---	----------------

Figure 8a (Prior Art)

U.S. Patent

Apr. 6, 2010

Sheet 15 of 22

US 7,693,956 B2

Top of Form

## Round-Trip Tickets. One-Way Prices.

Save up to 40% or more on round-trip tickets to virtually anywhere you want to fly. Priceline works directly with the major airlines to help you get great online deals.

**Name Your Own Price - Airfare**

Departure City:

Arrival City:


Departure Date:

Return Date:

Tickets:

**Air-Hotel**  
Click Here

**Travel Insurance**  
 Protect Your Purchase!  
Click Here






Figure 8b (Prior Art)

**CALIFORNIA DREAMIN' ON SALE WITH  
\$30 OFF PER PERSON FROM SOUTHWEST AIRLINES VACATIONS**

DALLAS, March 5, 2002 – Southwest Airlines Vacations Customers dreaming of a California vacation with movie stars, warm beaches, or cable cars can discover it for less with \$30 off per person to Los Angeles/Anaheim, San Diego, or the San Francisco Bay area. The California Dreamin' sale is available to Customers traveling from Phoenix and Tucson, Ariz; Kansas City, Mo; New Orleans; Las Vegas; Reno, Nev.; Albuquerque, N.M.; El Paso, Austin, Houston and San Antonio, Texas; Salt Lake City; Seattle; Portland, Ore. and all applicable California origins. Getaways to the Golden State are available on complete air/hotel vacations when purchased by May 24, 2002 for travel March 31 through June 29, 2002.

Customers can enjoy more value when they customize their complete vacation by purchasing the following California attraction passes:

- Purchase a one-day adult SeaWorld® San Diego admission and receive a second day admission free with San Diego and Los Angeles/Anaheim packages.
- Purchase a one-day adult Universal Studios Hollywood® admission and receive a second day admission free.

---

**CALIFORNIA DREAMIN' SALE COMPLETE VACATION PRICES**

Figure 8c (Prior Art)

# Horizon Air Lowers Fares from Fresno to Portland and Seattle

3/19/2002 2:57 p.m.

"These new fares dramatically reduce the cost for business travelers and provide a new lower regular fare for leisure travelers," said Patrick Zachwieja, vice president of marketing and planning. "These fares are in addition to any limited-time specials and sale fares that may appear from time to time."

A one-way fare with a 21-day advance from Fresno to Portland or Seattle is \$105, based on a roundtrip purchase. A 14-day advance fare is \$135, based on a roundtrip purchase.

The above fares include Federal Excise Tax and any applicable fuel surcharges, but they do not include Passenger Facility Charges of up to \$18 (amount depends on itinerary), Federal Segment Tax of \$3 per segment (takeoff and landing) on itinerary, and September 11th Security Fee of \$2.50 per enplanement (up to \$10). Roundtrip sale fares may be limited and are non-refundable; however, the flight ticket can be charged for a \$50 fee, plus any applicable changes in fare. All fares are subject to change without notice.

Horizon Air has served Fresno since March of 1999. Horizon Air serves 41 cities throughout Washington, Oregon, Idaho, Montana, California, Arizona, British Columbia, and Alberta.

In the last few months, Horizon Air has added service in several cities: between Medford and Los Angeles, from Boise and Palm Springs to Sacramento, and from Portland and San Jose to Tucson. Horizon will add service from San Diego and San Francisco to Boise on April 1, and from Portland and Boise to Denver on April 28.

Figure 8d (Prior Art)

U.S. Patent

Apr. 6, 2010

Sheet 18 of 22

US 7,693,956 B2

**Southwest Airlines Schedule Request**

**Where are you traveling?**

Top of Form

Origin	Destination	Month	Day
Reno, NV - RNO	Reno, NV - RNO	March	13
Sacramento, CA - SMF	Reno, NV - RNO	April	14
Salt Lake City, UT - SLC	Reno, NV - RNO	May	15
San Antonio, TX - SAT	Reno, NV - RNO	June	16
San Diego, CA - SAN	Reno, NV - RNO	July	17
San Jose, CA - SJC	Reno, NV - RNO	August	18
	Reno, NV - RNO	September	19

Bottom of Form

Display Schedule

Seattle/Tacoma WA - SEA

San Jose, CA - SJC

Figure 9a

U.S. Patent

Apr. 6, 2010

Sheet 19 of 22

US 7,693,956 B2

Top of Form									
<b>Shop for Flights and Fares</b>									
<a href="#">Look up a Saved Reservation</a>									
<b>From</b>		<b>To</b>		<b>Departing:</b>		<b>Returning:</b>			
Seattle		San Jose		Mar 19		Mar 18		7 am	
Seattle		San Jose		Mar 19		Mar 18		7 am	
<b>Adults:</b>		<b>Children:</b>		<b>Ages</b>		<b>Ages</b>			
1		0		13+		2-12			
Anytime age 2-17, traveling alone?									
<b>Cabin/Fare: About Fare Options</b>									
Coach/Restricted Fare									
<b>ONE WAY</b>		<b>ROUND TRIP</b>							
6		6							
9		7							
N		FitFares							
<b>Multiple Destinations</b>									
Bottom of Form									

Figure 9b

**U.S. Patent**

Apr. 6, 2010

Sheet 20 of 22

**US 7,693,956 B2**1002a


1. www.GeneralTravel.com

From:  To:   
 Departure date:     
 Return date:

2. United Airlines

From:  To:   
 Departure date:     
 Return date:

3. SouthWest Airlines

From:  To:   
 Departure date:     
 Return date:

4. FareSeek Flights -- Discount Airfare. Discount Flights Hotels - ...

20% to 30% Discount on International Flights! Departure. From: To:  
 www.fareseek.com/ - 83k - Mar 23, 2003 -

5. AirTech Specials

... but your chances of getting bumped off any of these flights are the same as ... 3/22  
 Maui to San Francisco \$129 one way, 3/22 Baltimore to Copenhagen \$189 one way, ...  
 www.airtech.com/specials.html - 30k - Mar 23, 2003 -

**Figure 10a**

U.S. Patent

Apr. 6, 2010

Sheet 21 of 22

US 7,693,956 B2

1002b

flights from baltimore to san francisco

[www.GeneralTravel.com](http://www.GeneralTravel.com)

From:  To:   
 Departure date:     
 Return date:

1. FareSeek Flights -- Discount Airfare. Discount Flights Hotels - ...  
 20% to 30% Discount on International Flights! Departure. From: To:  
[www.fareseek.com/](http://www.fareseek.com/) - 83k - Mar 23, 2003 -
2. AirTech Specials  
 ... but your chances of getting bumped off any of these flights are the same as ... 3/22  
 Maui to San Francisco \$129 one way, 3/22 Baltimore to Copenhagen \$189 one way, ...  
[www.airtech.com/specials.html](http://www.airtech.com/specials.html) - 30k - Mar 23, 2003 -
3. Discount Hotels: Hotels and Car Rentals from Travel Guys  
 ... FLIGHTS Discount Airfare Cheap Flights HOT DISCOUNTS ... Anaheim Atlanta Atlantic City  
 Baltimore Biloxi Boston ... San Antonio San Diego San Francisco Seattle Tampa ...  
[www.travelguys.net/](http://www.travelguys.net/) - 15k - Mar 23, 2003 -
4. USATODAY.com - Flight cuts make 1-day business trips difficult  
 ... says US Airways has cut so many flights between Baltimore and the West Coast, he  
 often has to stay an extra night on business trips to San Francisco and Los ...  
[www.usatoday.com/money/biztravel/2002-04-30-late-flight.htm](http://www.usatoday.com/money/biztravel/2002-04-30-late-flight.htm) - 27k - Mar 23, 2003 -
5. Courier Travel Standby Member Page  
 ... These listings only show flights where over 50 seats are ... Sold Out This Week, Baltimore,  
 MD - US, Cancun, Mexico, ... 2,3,4, Honolulu, Hawaii -US, San Francisco, CA - US, ...  
[www.couriertravel.org/standby.asp](http://www.couriertravel.org/standby.asp) - 30k -

Figure 10b



U.S. Patent

Apr. 6, 2010

Sheet 22 of 22

US 7,693,956 B2

1002c

1. [FareSeek Flights -- Discount Airfare. Discount Flights Hotels - ...](#)  
20% to 50% Discount on International Flights! Departure. From: To:  
[www.fareseek.com/](#) - 83k - Mar 23, 2003 -
2. [AirTech Specials](#)  
... but your chances of getting bumped off any of these flights are the same as ... 3/22  
Mau to San Francisco \$129 one way, 3/22 Baltimore to Copenhagen \$189 one way, ...  
[www.airtech.com/specials.html](#) - 30k - Mar 23, 2003 -
3. [Discount Hotels: Hotels and Car Rentals from Travel Guys](#)  
... FLIGHTS Discount Airfare Cheap Flights HOT DISCOUNTS ... Anaheim Atlanta Atlantic  
City  
Baltimore Biloxi Boston ... San Antonio San Diego San Francisco Seattle Tampa ...  
[www.travelguys.net/](#) - 15k - Mar 23, 2003 -
4. [USATODAY.com - Flight cuts make 1-day business trips difficult](#)  
... says US Airways has cut so many flights between Baltimore and the West Coast, he  
often has to stay an extra night on business trips to San Francisco and Los ...  
[www.usatoday.com/money/biztravel/ 2003-04-30-late-flight.htm](#) - 27k - Mar 23, 2003 -
5. [Courier Travel Standby Member Page](#)  
... These listings only show flights where over 50 seats are ... Sold Out This Week, Baltimore,  
MD - US, Cancun, Mexico, ... 2,3,4, Honolulu, Hawaii -US, San Francisco, CA - US, ...  
[www.couriertravel.org/standby.asp](#) - 30k -

[www.GeneralTravel.com](#)

From:

To:

Departure date:

Return date:

[United Airlines](#)

From:

To:

Departure date:

Return date:

Figure 10c

US 7,693,956 B2

1

## ACCESSING DEEP WEB INFORMATION USING A SEARCH ENGINE

### RELATED APPLICATION

The present invention is a non-provisional application of provisional application No. 60/366,817 filed on Mar. 25, 2002, entitled "A Method of Querying Multiple Deep Web Sites Using a Search Engine", claims priority to said provisional application, and incorporates its specification in its entirety by reference.

### FIELD OF THE INVENTION

The present invention relates to the field of data processing. More specifically, the present invention is related to accessing deep web pages using a search engine.

### BACKGROUND OF THE INVENTION

Currently, Internet search engines index the information found on static web pages on public websites. Static web pages are web pages that exist before and after a user accesses them. A search engine saves two pieces of information from each static web page it indexes. The first piece of information is the location of the web page, commonly referred to as the uniform resource locator (URL). The second piece of information is one or more keywords from the text of the web page. A search engine is then, essentially, a collection of URLs and associated keywords.

Deep web information, however, is not contained on static web pages. Instead, it is produced in response to a specific query by a user and results in dynamic web pages. Search engines themselves are deep websites. A dynamic web page only exists after a user's query and disappears after the user exits the dynamic web page. As a result, deep web information or the dynamic web pages that can be generated therefrom are not typically available from search engines.

As a result, a user using a search engine to locate information, for certain type of information searching, often finds himself/herself not finding the right information, or having to re-enter the search criteria, as query parameters of queries to generate the dynamic web pages having the desired information.

FIGS. 8a-8d illustrate one example of this problem. For the illustrated example, a user uses a search service/engine (such as Goggle) to search for "Flights from Seattle to San Jose". FIG. 8a illustrates selected portions of an example search answer web page 800 returned from the search service/engine. Example search answer web page 800 includes indices 802 and 804 indexing flight deals on Priceline and Orbitz's websites respectively, and indices 806 and 808 indexing static promotional pages on Southwest and Alaska Airline's websites respectively.

FIG. 8b illustrates the example static web page indexed by index 802. It is essentially a form for facilitating a user to search for specific flight deals available on Priceline's website. In other words, because the flight deal web page for flights between Seattle and San Jose, if available, is dynamically generated based on flight deal information in a database, the proper query answer web page is not indexed by the search service/engine. Consequently it is not available. Only the generic static root web page (which could have been retrieved with merely the search criteria of "flight deals") is indexed, resulting in the user essentially having to start the search over again, once he/she enters Priceline's website.

2

Similar disadvantages exist, if the Orbitz website is entered via index 804.

FIGS. 8c-8d illustrate the static promotional web pages indexed by indices 806-808. They are merely promotional web pages, and not the flight schedule or availability web pages that the user is looking for. Again, the reason is that the flight schedule and availability web pages are dynamically generated based on the flight schedule and availability information in the databases of Southwest and Alaska Airlines.

FIGS. 9a-9b illustrate the end results that the user desires. That is, in response to the user's provision to a search engine, of the search criteria "flights from Seattle and San Jose", the user is provided with a search answer page having indices that index not only the relevant static web pages, but also the relevant dynamic web pages to be generated, such as would be provided using this invention. Further, on selection of one of such indices, the user can be taken to the indexed dynamic web page, with the search criteria preserved, i.e. the from and to cities of Seattle and San Jose, filled in, without requiring the user to re-fill or re-select the search criteria again.

### BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will be described by way of exemplary embodiments, but not limitations, illustrated in the accompanying drawings in which like references denote similar elements, and in which:

FIG. 1 illustrates network view of a computing environment suitable for practicing present invention, in accordance with one embodiment;

FIG. 2 illustrates an overview of the method of the present invention, in accordance with one embodiment;

FIGS. 3a-3b illustrate a number of reconciled queries with their associated query parameters, of a number of websites, for a number of example search criteria, created in accordance with the teachings of the present invention;

FIGS. 4a-4b illustrate a number of example proxy web page specifications, created in accordance with the teachings of the present invention;

FIGS. 5a-5b illustrate two example indices created using proxy web page specifications of the present invention;

FIGS. 6a-6d illustrate a source view and an end-user view each of two example pre-filled forms for two example queries employing the POST method to post its associated query parameters to the corresponding websites;

FIG. 7 illustrates an example computer system suitable for use as either a client or to host a proxy function, a search engine or a web server, to practice the present invention, in accordance with one embodiment;

FIGS. 8a-8d illustrate the inadequacies of the prior art in facilitating a user in accessing deep web information using a search engine, for an example information search;

FIGS. 9a-9b illustrate two user desired results of the example information search of FIG. 8a-8d; and

FIGS. 10a-10c illustrate three example search result pages, in accordance with an alternate embodiment of the present invention.

### DETAILED DESCRIPTION OF EMBODIMENTS OF THE INVENTION

The present invention includes search engines enhanced to effectively index dynamic web pages, companion proxy servers, and related methods and apparatuses.

Part of the description will be presented in terms commonly employed by those skilled in the art to convey the substance of their work to others skilled in the art, including

## US 7,693,956 B2

3

but are not limited to terms such as websites, static and dynamic web pages, queries, query parameters, query answer pages, search criteria, search answer web pages and forms. On a computing device, these logical entities are represented and manipulated in the form of electrical and/or optical signals.

In the following description, various aspects of the present invention will be described. However, it will be apparent to those skilled in the art that the present invention may be practiced with only some or all aspects of the present invention. For purposes of explanation, specific numbers, materials and configurations are set forth in order to provide a thorough understanding of the present invention. However, it will be apparent to one skilled in the art that the present invention may be practiced without the specific details. In other instances, well-known features are omitted or simplified in order not to obscure the present invention.

Various operations will be described as multiple discrete steps in turn, in a manner that is most helpful in understanding the present invention, however, the order of description should not be construed as to imply that these operations are necessarily order dependent. In particular, these operations need not be performed in the order of presentation.

The use of heading and sections are to facilitate understanding of the present invention, and they are not to be limiting on the present invention. The phrase “in one embodiment” is used repeatedly. The phrase generally does not refer to the same embodiment, however, it may. The terms “comprising”, “having” and “including” are synonymous, unless the context dictates otherwise.

## Overview

Referring now to FIG. 1, wherein a network view of a computing environment suitable for practicing the present invention, is shown. As illustrated, computing environment **100** includes a number of websites **108** offering services and/or information. In particular, each website **108** includes dynamic web page generator **118** to generate various dynamic web pages to facilitate delivery of the services and/or information offered.

Typically, dynamic web page generator **118** generates the various dynamic web pages using information stored in various databases (not shown) of websites **108**, which may be co-resident or remotely disposed from the server(s) hosting websites **108** (but communicatively coupled). Each dynamic web page is typically generated in response to a query having a particular set of associated query parameters. Accordingly, these dynamic web pages will also be referred to as query answer web pages.

Further, the query answer web pages may be generated using Common Gateway Interface (CGI) scripts, in conjunction with instructions of one or more programming languages (Perl, C, C++, Basic, Fortran, Pascal, etc.), or a proprietary dynamic web page generator, such as Microsoft's generated ASP pages.

For the illustrated embodiment, communication with websites **108** are conducted using the Hypertext Transmission Protocol (HTTP) over Transmission Control Protocol/Internet Protocol (TCP/IP), and the dynamically generated web pages are specified using the Hypertext Markup Language (HTML). For some queries, the associated query parameters are submitted to websites **108** using a GET method, i.e. the associated query parameters of the queries are submitted in conjunction with the corresponding URLs, identifying the locations of the corresponding websites **108**. Whereas, for other queries, the associated query parameters are submitted

4

to websites **108** using a POST method, i.e. the associated query parameters of the queries are submitted separately (typically subsequently), after corresponding connections have been established with the websites **108**, using their corresponding URLs. Thus, in the second method called the POST method, the URL is disassociated from the input parameters. The associated input parameters with the URL would make the resulting URL unique. However, without a unique URL, that is, without the association of the FORM URL with the specific input parameters, the completed FORM cannot be indexed by a search engine or cannot even be bookmarked by a browser.

In alternate embodiments, the present invention may be practiced with communication with websites **108** being practiced using other communication protocols, including but are not limited to HTTP Secure (HTTPS), File Transfer Protocol (FTP), Secure FTP (SFTP), Telnet, Secure Shell, Secure Copy (SCP), Goopher, and so forth.

Continuing to refer to FIG. 1, computing environment **100** further includes a number of clients **102**, and search services **104**, coupled to each other and to websites **108** via network **110**. Clients **102**, assisted by search services **104**, selectively consume the services and/or information offered by websites **108**.

For the embodiment, clients **102** include browsers **112**, and use browsers **112** to consume the service and/or information offered by websites **108** in the form of web pages. Clients **102** further include operating system services (not shown) in support of browsers **112**. Browsers **112**, supported by the operating system services, are equipped to communicate with search services **104** and websites **108** in one or more of the earlier described protocols.

In alternate embodiments, browsers **112** and the underlying operating system services may be fully integrated. In other embodiments, some functions of browsers **112** and the underlying operating system services may be offered by an intermediate “middle” layer of services instead.

For the embodiment, search services **104** include search engine **114**, web page indices **115** indexing web pages of websites **108** offering services and/or information, and index generator **113** to generate indices **115**. As will be described in more detail below, under the present invention, web page indices **115** also advantageously index web pages of websites **108** that are dynamically generated in response to queries having associated query parameters.

Similar to clients **102**, search services **104** further include operating system services (not shown) in support of elements **113-115**. Search engine **114** supported by the operating system services, are equipped to communicate with clients **102** and websites **108** in the appropriate protocols.

In alternate embodiments, search engine **114** and the underlying operating system services may be fully integrated. In other embodiments, some functions of search engine **114** and the underlying operating system services may be offered by an intermediate “middle” layer of services instead.

For the embodiment, search services **104** are complemented by at least one proxy server **106**, which is coupled to websites **108**, clients **102**, and search services **104** via network **110**. Proxy server **106** includes in particular proxy function **116**, and associated query parameter values **117** for pre-filling query forms to be provided to their websites **108**. More specifically, for the earlier described HTML and HTTP oriented embodiment, pre-filled query forms facilitate posting of the associated query parameters to their websites **108** via the POST method.

The term “form” as used herein in the specification and in the claims, in general, refers to an electronic form which may

## US 7,693,956 B2

5

be one of any one of a number of form types, including but are not limited to, HTML FORMs, Windows FORMs, ASP generated FORMs, Java FORMs, Javascript FORMs, directX FORMs, and so forth, as well as form-like documents and/or data structures/organizations, unless the context of a particular usage instance specifically narrows the meaning for the particular usage instance to a particular form type.

As illustrated, indices **115** effectively index the dynamically generated query answer web pages, and if applicable, associated query parameter values **117** for pre-filling query forms are generated in accordance with proxy web page specifications **119** of the dynamically generated query answer web pages.

Except for indices **115** that effectively index the dynamically generated query answer web pages, associated query parameter values **117** for pre-filling query forms (if applicable) to facilitate submission of associated query parameters of queries of websites **108**, and proxy web page specifications **119**, clients **102** (including browsers **112**), search services **104** (including search engine **114** and indices **115** that index static web pages), proxy server **106** (including proxy function **116**), websites **108** (including dynamic web page generators **118** and their databases (not shown)), and network **110** represent a broad range of these respective elements known in the art or to be designed. In particular, network **110** may include one or more private and/or public networks, such as the Internet. Their functions and constitutions are known, accordingly will not be further described.

Otherwise, the method of the present invention, proxy web page specifications **119**, indices **115** that effectively index the dynamically generated query answer web pages, and associated query parameter values **117** for pre-filling forms (if applicable) to facilitate submission of associated query parameters of queries of websites **108**, will be described in turn, in the description to follow.

## Initialization/Set up

FIG. **2** illustrates an overview of a method of the present invention, in accordance with one embodiment. More specifically, FIG. **2** illustrates the initialization or set up operations for setting up search services **104**, and if applicable, proxy servers **106**, to facilitate clients **102** to access deep web information of websites **108**.

As illustrated, for each set of search criteria of interest, block **202**, the queries including the associated query parameters, and the manner they are submitted to websites **108**, e.g. whether through the GET method or the POST method, are identified, block **204**. The identified queries are then reconciled, block **206**.

The above operations are best illustrated by a couple of examples. The first example illustrates how queries of various used car websites of interest for various used car searches of a HTML/HTTP based embodiment are identified and reconciled, and the second example illustrates how queries of various airline websites of interest for various used flight schedule/availability searches of a HTML/HTTP based embodiment are identified and reconciled.

## Used Car Sites

Suppose two used car deep web sites **108** are of interest. These sites are Automall.com and JoesCars.com, and for ease of understanding, their dynamically generated web pages are assumed to be HTML based. Moreover, communication with Automall.com and JoesCars.com are assumed to be HTTP based.

Further, the queries of their query answer web pages have similar query parameters. More specifically, Automall.com's

6

query parameters are car make, car model, and user's zip code. JoesCars.com's query parameters are car make, car model, user's zip code and car year or years.

The first step in identifying and reconciling the queries of these sites is to find a minimum number of query parameters for this group of queries. In this example case, JoesCars.com has one more query parameter, which is the year or years. Since Automall.com does not have this parameter, it can be eliminated. If JoesCars.com requires the year or years, the largest range of years may be selected. For example, if JoesCars.com provides support for years between 1982 and 2002, then the year range for its query will always be 1982-2002.

Choosing a default constant value for the car year leaves three query parameters for both sites to be specified. This, however, may be reduced further. If it is assumed that every model name is unique, then the user may be able to make a query with just two pieces of information. These are the car model and user's zip code. If the car make is required by both sites, this will be provided in the query. The user will not, however, be required to enter it in the search engine query.

In various embodiments, during this stage of query identification and reconciliation, the user's options for query parameters may also be expanded. For example, if a list exists which associates a zip code to a city and state, the process may make plans to allow for a user to substitute a city or state in the search engine query. The query that is sent to each deep web site **108** will contain the zip code even though the user entered a city or state. If the list of associated cities and states for zip codes is used, the minimum number of query parameters for used car sites is still two. However, the parameters are now car model and user location.

Once the minimum number of query parameters is found, the next step is to identify how the query of each deep web site **108** in the group is invoked. This involves determining the query syntax and method of the HTTP request and the number and type of parameters required. Most of this information can be obtained by examining the HTML within the FORM statement of the dynamically generated query answer page. Any additional information that is required may be found by executing some sample queries.

Suppose that in the example given, it is found that Automall.com uses a GET HTTP request method and JoesCars.com uses a POST request method. Furthermore, Automall.com's URL syntax is

`http://www.automall.com/  
used.html?make=chevrolet&model=corvette&zip=22043`

and JoesCars.com's POST parameters are:

Make, Model, Year and Zip

The request method and query syntax for the deep web sites **108** are then known. What remains to be understood is how the GET and POST parameters need to be added. For example, if the make parameter for Automall.com is entered through a text field, then it may be insensitive to the case of a make name. However, if the make parameter is obtained from a pulldown, there is only one way of representing each make name.

Assume that Automall.com and JoesCars.com list car make and model in user pulldowns which translate to HTML SELECT statements. All of the possible combinations of make and model can then be obtained from the HTML. Assume that the Year input for JoesCars.com is a TEXT field and "1982-2002" is an acceptable range syntax. Finally, assume that both sites accept a five-digit zip code through a TEXT field. The method for invoking the deep web sites is then well understood.

US 7,693,956 B2

7

After defining the minimum set of query parameters and understanding the methods of invocation of each deep web site **108**, the next step is to develop a list of site queries and parameter values.

For example, an excerpt from Automall.com's list is illustrated in FIG. **3a**. The list contains the query URL followed by a set of query parameter values that a user might enter as a set of search criteria.

Note that the query URL contains the car make, but the set of query values does not. This is because the minimum set of query parameters was chosen. Note also that some query values contain the city name instead of the zip code, but the zip code was entered in the URL query. This is an example of the expansion of the possible query parameter values by using an additional file that associates zip codes and city names.

An excerpt from JoesCars.com list of site queries and parameter values **302-304** is also shown in FIG. **3a**.

The first noticeable difference between Automall.com's list and JoesCars.com's list is the URL destination. Instead of JoesCars.com's URL, the destination is a proxy server's URL (outside.com). This is because of the HTTP request method used on the JoesCars.com site. A search engine cannot execute a POST request from one of its links. As a result, the link given to the search engine must be a GET request to a third party proxy site that can execute the post and return the JoesCars.com results page to the user.

Another difference is the additional Year parameter. As mentioned above, all queries are given the default range of "1982-2002" for this additional parameter. Also, since each site might span a different range of years the list of possible query parameter values might be different. For example, JoesCars.com might show Ford Granada while this model might be too old for Automall.com. Remember that in this example the car models were found by searching the SELECT statements of the HTML.

Note that while these lists might contain all possible inputs for the minimum set of query parameters, not all items in the list may be submitted to the search engine. Instead, each item on the list might be compared to previous search engine queries for relevance. If such a query is never entered, it may be excluded from the information given to the search engine.

#### Airline Reservation Sites

Suppose that two deep web airline reservations sites **108** are of interest for various flight schedule and availability searches. One is a general travel site that lists the reservations of many different airlines. This site is called GeneralTravel.com. The other is a site for a specific airline. This site is called FastJet.com.

Again, for ease of understanding, the dynamically generated web pages of GeneralTravel.com and FastJet.com are assumed to be HTML based, and communications with the two sites are HTTP based.

As with the used cars, the first step in identifying and reconciling the queries of the websites is to identify the minimum set of query parameters. For airline reservations the minimum set of parameters is four. These are the "departure date", "from location", "return date" and "to location". These parameters may also be used for hotel and car rental reservations. However, the "from location" is not needed for these types of reservations.

Suppose GeneralTravel.com also requires the number of people traveling. Since this is not part of the minimum set of parameters, the default can be used. In this case the default is one person.

The next step is to understand the method of query invocation on each site. Suppose that both GeneralTravel.com and

8

FastJet.com use the GET HTTP request method. Also suppose that the URL syntax for each of these sites is as follows.

```
http://www.generaltravel.com/
res.html?persons=1&from=bwi&to=sfo&leav
e=0504&return=0506
```

```
http://www.fastjet.com/
res.html?fdest=bwi&tdest=sfo&start=0504&end=0506
```

The third step in the process is then to create the lists of site queries and parameter values. In the case of used car sites, the car models were limited by the values placed in the HTML select statements. Even the number of user zip codes was a finite set. However, the number of possible departure dates and return dates is infinite. As a result, the list of possible parameter values must be limited by some other means.

Suppose that most reservations are booked no more than one year in advance. Also, suppose that most travel takes place within one month. The possible parameter values are now considerably limited. First of all, the departure dates are then the next 365 days. For each departure date, there will be 30 return dates covering the following 30 days. Thus, there will be a total of 365 times 30, or 10,950 possible combinations of dates.

To complete the list of values, a list of all possible destinations must be available. Every possible combination of to and from destination can then be generated. Furthermore, the destinations can be specified as cities or airport codes. The complete list of parameter values is then 10,950 possible combinations of dates for each possible combination of to and from destinations.

An excerpt from GeneralTravel.com's list of site queries and parameter values **312-314** is also shown in FIG. **3b**.

Note that the URL contains information not in the query. In this case, this is the "number of persons traveling." Also, note that query may contain the city or the airport, but the URL only contains the airport code. This implies that a list correlating airport codes and city names was used to generate the list of parameter values.

Referring back to FIG. **2**, once the queries of the websites **108** of interest have been identified and reconciled, corresponding proxy web page specifications **119** for the dynamically generated query answer web pages are created, block **208**, and submitted to search services **104** for processing to create indices **115** to index the dynamically generated query answer web pages.

More specifically, as earlier described, corresponding proxy web page specifications **119** are submitted to index generator **113**, which generates, in response, the earlier described indices **115**.

As described earlier, index generator **113** may be any one of a number of such elements known in the art or to be designed. Its functions and constitution are known, and except for its usage, the manner it is constituted is not essential to the present invention.

FIGS. **4a-4b** illustrate one each, an example proxy web page specification **119a-119b** for the queries of Automall.com and Generaltravel.com. In alternate embodiments, the information may be provided to search services **104** via other formats and/or approaches.

FIGS. **5a-5b** illustrate the corresponding example indices that may be generated to effectively index the dynamic query answer pages for the earlier described search examples, i.e. used car search, and flight schedule and availability search. As illustrated, example indices **500a** may comprise indices **502a-502b** that advantageously index the dynamically generated query answer pages for the example used car search, and example indices **500b** may comprise indices **502c-502b** that advantageously index the dynamically generated query

## US 7,693,956 B2

9

answer pages for the example flight schedule/availability search, and thereby overcoming the prior art deficiency of not having any index that indexes dynamically generated web pages.

Referring again to FIG. 2, and recalling from earlier descriptions, for queries of the HTML and HTTP embodiment employing a FORM and the POST method to post their associated query parameters to their websites 108, associated query parameter values 117 for pre-filling query forms are created and pre-stored in proxy server 106 to facilitate the form filing and parameter value posting. Accordingly, upon creation of appropriate proxy web page specifications 119 and their provisions to index generator 113 to generate indices 115, proxy web page specifications 119 are employed to create the associated query parameter values 117 for pre-filling forms and pre-store them with proxy server 106 for use during operation.

Associating query parameter values 117 for pre-filling query forms may be stored on proxy servers 106 using any one of a number of data structures and/or organizations known in the art or to be designed, including but are not limited to hash tables, flat files, relational databases, and so forth.

FIGS. 6a and 6c illustrate one each, a source view of an example pre-filled form 117a for the above described used car search example, and a source view of an example pre-filled form 117b for the above described flight reservation example, to facilitate posting of associated query parameters of queries corresponding to two dynamically generated query answer pages.

As illustrated, pre-filled form 117a has the values “Ford” and “Focus” pre-selected for the Make and Model parameters. Further, pre-filled form 117a has the zip code value “22043” pre-set for the zip code parameter.

Similarly, pre-filled form 117b has the parameters values for the departing city, destination city, number of passenger, whether it is a round trip reservation, departure date, return date, and so forth pre-filled.

For ease of understanding, some of the less illustrative source codes (in the “blank” areas) are omitted.

FIGS. 6b and 6d illustrate the end user view of the corresponding example pre-filled forms 117a and 117b for the above described used car and flight reservation search examples, when a user selects the index that effectively indexes the corresponding dynamically generated query answer page.

Those skilled in the art will appreciate that the examples have been purposely kept simple for ease of understanding. Nevertheless, they convey the substances of the present invention, even though in practice, typically more query parameters and query values are involved.

## Operation

Thus, during operation, on receipt of a set of search criteria from a client 102, search service 104 is able to return a search answer page comprising indices indexing query answer web pages that are not persistent, and are to be dynamically generated.

As described earlier, example search result web pages for the earlier described examples, i.e. used car search and flight schedule and availability search may comprise indices 502a-502b and 502c-502d that advantageously index dynamically generated web pages, thereby overcoming the prior art deficiency of not having any index that indexes dynamically generated web pages.

Further, on selection, the search criteria will be carried forward, and incorporated into the dynamically generated

10

answer web page, thereby also overcoming the prior art disadvantage of requiring the user to re-enter the search criteria. As described earlier, this is effectuated through the use of proxy server 106 and the complementary associated query parameter values 117 for pre-filling query forms pre-generated and pre-stored on proxy server 106.

In alternate embodiments, in lieu of retrieving the relevant forms in real time from the applicable websites 108, the present invention may be practiced with the forms pre-fetched, pre-filled and cached.

For the illustrated embodiment, preservation and transfer of the search criteria or query parameters are accomplished without resorting to the use of cookies (and client approval of such use). However, in alternate embodiments, some or all of the query parameters may be preserved and/or transferred involving the use of cookies (and client approval of such use).

Moreover, the present invention may be practiced with the preservation and transfer of other state information, in addition to the search criteria. Furthermore, the present invention may also be practiced in conjunction with the employment of one or more security techniques.

## Example Computer System

FIG. 7 illustrates an example computing system suitable for use as either a client or to host the proxy function, the search engine or a web server to practice the present invention. As a client system, computing system 700 may be a desktop computer system, a laptop computer system, a tablet computing device, a palm sized computing device, a wireless mobile phone, a set-top box, an Internet appliance and the like. As a server, computing system 700 may be a single or a cluster of computer systems.

As shown, exemplary computing system 700 includes one or more processors 702 and system memory 704. Additionally, system 700 may include mass storage devices 706 (such as diskette, hard drive, CDROM and so forth), input/output devices 708 (such as keyboard, cursor control and so forth) and communication interfaces 710 (such as network interface cards, modems and so forth). The elements are coupled to each other via system bus 712, which represents one or more buses. In the case of multiple buses, the buses are bridged by one or more bus bridges (not shown).

Each of these elements performs its conventional functions known in the art. In particular, system memory 704 and mass storage 706 are employed to store a working copy and a permanent copy of the programming instructions and/or data, implementing the teachings of the present invention (e.g. the enhanced indices of the search engine, the complementary pre-filled forms, and so forth). The programming instructions may be instructions of any one of a number of known or to be designed languages, including but are not limited to C, C++, Perl, Java, Javascript, XML, HTML and so forth. All or a portion of the permanent copy of the programming instructions/data may be loaded into mass storage 706 in the factory, in the field (via a distribution medium) or through communication interface 710 (from a distribution server (not shown)).

The constitution of these elements 702-712 are known, and accordingly will not be further described.

## Alternate Embodiments

While the present invention has been described in terms of the foregoing embodiments, those skilled in the art will recognize that the invention is not limited to the embodiments described. The present invention can be practiced with modification and alteration within the spirit and scope of the appended claims.

## US 7,693,956 B2

11

In particular, while for ease of understanding, the search services and the proxy server have been described as two separate entities, in alternate embodiments, the two may be combined, or one or more functions of one or more of search services and proxy server may be distributed among additional implementation entities.

In one alternate embodiment, the search services and the proxy server may be effectively combined, by having the search services automatically select the indices that index the dynamically generated web pages, and cause the query result web pages to be embedded as part of the search result page. The query results may be embedded into the search result page as other “hits” (as example search result page **1002a** illustrated in FIG. **10a**), or highlighted, e.g. placed before the normal “hits” (as example search result page **1002b** in FIG. **10b**) or placed on the side of the normal “hits” (as example search result page **1002c** illustrated in FIG. **10c**).

Further, in alternate embodiments, the present invention may also be suitably implemented, at least in part, in firmware or hardware, or some suitable combination of at least two of the three. Such firmware or hardware embodiments may include programmable logic devices, such as programmable array logic (PALs) and programmable logic arrays (PLAs), field programmable gate arrays (FPGAs), application specific integrated circuits (ASICs), large scale integrated circuits (LSIs), very large scale integrated circuits (VLSI) or the like—to form the various types of modules, circuitry, controllers, routines and systems described and claimed herein.

## Conclusion and Epilogue

Thus, it can be seen from the above descriptions, a novel method, including associated apparatuses, for accessing deep web information has been described. The description is to be regarded as illustrative instead of restrictive on the present invention, which scope is set forth in the claims to follow.

What is claimed is:

1. A computer implemented method, comprising:

receiving by a search service computing server on the Internet, from a client device having access to the Internet, a search request, wherein the search request includes one or more search criteria and directs the search service computing server to search the Internet for a plurality of information locations having information associated with the plurality of search criteria; and

returning to the client device, in response to the search request, by the search service computing server, an answer page having a plurality of answers identifying a plurality of information locations on the Internet having information associated with the one or more search criteria, wherein at least a first of the answers associated with a first information page of a first information location having information associated with the search criteria does not include any input field displayed on the answer page, and at least a second of the answers associated with a second information page of a second information location having information associated with the search criteria includes at least one input field displayed on the answer page, the input field being associated with the second information page;

wherein the second answer is placed in the answer page either ahead of or at a side of the first answer;

wherein the second answer including at least one input field further includes an index indexing to the second information location from which the input field is generated, the index including the one or more search criteria and a set of one or more associated parameters, and

12

the set of one or more associated parameters including at least one parameter variable corresponding to the at least one input field.

2. The method of claim 1, wherein the second answer including at least one input field includes a plurality of input fields associated with the second information page, organized in one or more rows.

3. The method of claim 1, wherein the method further comprises transmitting a query to the second information location with whose second information page the input field is associated, the query including one or more query parameters, and the one or more query parameters including at least one value input through the at least one input field.

4. A computer implemented method comprising:

receiving by a search service computing server on the Internet, from a client device having access to the Internet, a search request, wherein the search request includes a plurality of search criteria and directs the search service computing server to search the Internet for a plurality of information locations having information associated with the plurality of search criteria, the plurality of search criteria including a first and a second physical location name or identifier; and

returning to the client device, in response to the search request, by the search service computing server, an answer page having a plurality of answers identifying a plurality of information locations having information potentially associated with the first physical location, the second physical location, or both, wherein at least a first of the answers associated with a first information page of a first information location having information associated with the search criteria does not include any input field displayed on the answer page, and at least a second of the answers associated with a second information page of a second information location having information associated with the search criteria includes at least one input field displayed on the answer page for entry of at least a flight departure date for a flight between the first and second physical locations, the at least one input field associated with the second information page;

wherein the second answer is placed in the answer page either ahead of or at a side of the first answer;

wherein the second answer including at least one input field further includes an index indexing to the second information location from which the input field is generated, the index including the one or more search criteria and a set of one or more associated parameters, and the set of one or more associated parameters including at least one parameter variable corresponding to the at least one input field.

5. The method of claim 4, wherein the at least a second answer including at least one input field, includes a plurality of input fields and configured to enable the user to further indicate a flight return date for flights between the first and second physical locations.

6. The method of claim 5, wherein the plurality of input fields are organized in one or more rows.

7. The method of claim 5, wherein the plurality of input fields are organized into two rows, one row for indicating the flight departure date and another row for indicating the flight return date.

8. The method of claim 4, wherein the method further comprises transmitting a query to the second information location from whose second information page the input field is associated, the query including one or more query param-

US 7,693,956 B2

**13**

eters, and the one or more query parameters including at least a flight departure date input through the at least one input field.

**9.** The method of claim **4**, wherein at least one of the first and second locations is a city.

**14**

**10.** The method of claim **4**, wherein at least one of the first and second locations is an airport.

\* \* \* \* \*



UNITED STATES PATENT AND TRADEMARK OFFICE  
Certificate

Patent No. 7,693,956 B2

Patented: April 6, 2010

On petition requesting issuance of a certificate for correction of inventorship pursuant to 35 U.S.C. 256, it has been found that the above identified patent, through error and without any deceptive intent, improperly sets forth the inventorship.

Accordingly, it is hereby certified that the correct inventorship of this patent is: Michael Z. Moricz, Bellevue, WA (US).

Signed and Sealed this Twenty-eighth day of December 2010.

ARIO ETIENNE  
*Supervisory Patent Examiner*  
Art Unit 2457  
Technology Center 2400